

DE 01093
#513
5/13/08

CERTIFIED COPY OF
PRIORITY DOCUMENT



**Prioritätsbescheinigung über die Einreichung
einer Patentanmeldung**

Aktenzeichen: 101 06 581.7

Anmeldetag: 13. Februar 2001

Anmelder/Inhaber: Philips Corporate Intellectual
Property GmbH, Hamburg/DE

Bezeichnung: Spracherkennungssystem, Trainingseinrichtung und
Verfahren zum iterativen Berechnen freier Parameter
eines Maximum-Entropie-Sprachmodells

IPC: G 10 L 15/06

Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der ur-
sprünglichen Unterlagen dieser Patentanmeldung.

München, den 02. November 2001
Deutsches Patent- und Markenamt
Der Präsident

Im Auftrag

Weihmayr



BESCHREIBUNG

Spracherkennungssystem, Trainingseinrichtung und Verfahren zum iterativen Berechnen freier Parameter eines Maximum-Entropie-Sprachmodells

- Die Erfindung betrifft ein Verfahren zum iterativen Berechnen freier Parameter $\lambda_a^{ortho(n)}$ eines Maximum-Entropie-Sprachmodells MESM in einem Spracherkennungssystem mit Hilfe des Generalised-Iterative-Scaling-Trainingsalgorithmus gemäß folgender Formel:

$$\lambda_a^{ortho(n+1)} = G(\lambda_a^{ortho(n)}, m_a^{ortho}, \dots) \quad (1)$$

10 wobei:

n : einen Iterationsparameter;
 G : eine mathematische Funktion;
 α : ein Merkmal in dem MESM; und

m_a^{ortho} : einen gewünschten orthogonalisierten Randwert in dem MESM für das
 15 Merkmal α bezeichnen.

Die Erfindung betrifft weiterhin ein im Stand der Technik bekanntes computerunterstütztes Spracherkennungssystem sowie eine bekannte computerunterstützte Trainingseinrichtung, in denen das beschriebene Verfahren eingesetzt wird.

20

Ausgangspunkt für die Erstellung eines Sprachmodells, wie es in einem computergestützten Spracherkennungssystem zur Erkennung von eingegebener Sprache verwandt wird, ist eine vordefinierte Trainingsaufgabe. Die Trainingsaufgabe bildet bestimmte statistische Muster in der Sprache eines zukünftigen Benutzers des Spracherkennungssystems in ein System
 25 von mathematisch formulierten Randbedingungen ab, welches im allgemeinen folgende Gestalt hat:

$$\sum_{(h,w)} N(h) \cdot p(w|h) \cdot f_\alpha(h,w) = m_\alpha \quad (2)$$

wobei:

$N(h)$: die Häufigkeit der Historie h in einem Trainingskorpus;

$P(w|h)$: Wahrscheinlichkeit $p(w|h)$, mit welcher sich ein vorgegebenes Wort w an eine bisherige Wortfolge h (Historie) anschließt;

$f_{\alpha}(h,w)$: eine binäre Merkmalsfunktion für ein Merkmal α ; und

m_{α} : einen gewünschten Randwert in dem System von Randbedingungen;

10 bezeichnet.

Die Lösung dieses Systems von Randbedingungen, d.h. der Trainingsaufgabe, wird durch das sogenannte Maximum Entropie Sprachmodell MESM gebildet, welches eine geeignete Lösung des Systems von Randbedingungen in Form einer geeigneten Definition der

15 Wahrscheinlichkeit $p(w|h)$ angibt, die wie folgt lautet:

$$p(w|h) = p_{\lambda}(w|h) = \frac{1}{Z_{\lambda}(h)} \cdot \exp \left(\sum_{\alpha} \lambda_{\alpha} \cdot f_{\alpha}(h,w) \right) \quad (3)$$

wobei:

20 $Z_{\lambda}(h)$: einen historienabhängigen Normierungsfaktor;

λ_{α} : einen freien Parameter zum Merkmal α ;

λ : den Satz aller Parameter

bezeichnen. Für die übrigen Parameter gilt ihre obige Definition.

25

Die binäre Merkmalsfunktion $f_{\alpha}(h,w)$ trifft beispielsweise eine binäre Entscheidung, ob vorgegebene Worte an bestimmten Stellen in vorgegebenen Wortfolgen h,w enthalten sind.

Ein Merkmal α kann im allgemeinen ein einzelnes Wort, eine Wortfolge, eine Wortklasse (z.B. Farbe oder Verben), eine Folge von Wortklassen oder komplexere Muster bezeichnen.

- 5 In Fig. 4 sind vorbestimmte Merkmale in einem Sprachmodell beispielhaft dargestellt. So repräsentieren die dargestellten Unigramme jeweils ein einzelnes Wort, die Bigramme jeweils eine Wortfolge bestehend aus zwei Worten und das abgebildete Trigramm eine Wortfolge bestehend aus drei Worten. Das Bigramm "OR A" schließt das Unigramm "A" ein und umfasst darüber hinaus ein weiteres Wort; es wird deshalb als höher-reichweitig
 10 gegenüber dem Unigramm "A" bezeichnet. Analog ist das Trigramm "A WHITE HOUSE" von höherer Reichweite als das Unigramm "HOUSE", oder als das Bigramm "WHITE HOUSE".

- Die freien Parameter λ werden so angepasst, dass Gleichung 3 eine Lösung für das System
 15 von Randbedingungen gemäß Gleichung 2 darstellt. Diese Anpassung erfolgt üblicherweise mit Hilfe bekannter Trainingsalgorithmen. Ein Beispiel für einen derartigen Trainingsalgorithmus ist der sogenannte Generalized Iterative Scaling GIS- Algorithmus, wie er z.B. in J.N. Darroch and D. Ratcliff, „Generalized iterative scaling for log linear models“, Annals Math. Stat., 43(5):1470-1480, 1972 beschrieben wird.

20

Dieser GIS-Algorithmus sieht eine iterative Berechnung der freien Parameter λ vor. Traditionell erfolgt diese Berechnung allerdings recht langsam. Zur Beschleunigung dieser Berechnung wird im Stand der Technik vorgeschlagen, die Merkmalsfunktionen $f_{\alpha}(h,w)$ in dem System von Randbedingungen gemäß Gleichung (2) durch orthogonalisierte

- 25 Merkmalsfunktionen $f_{\alpha}^{ortho}(h,w)$ zu substituieren; siehe dazu R. Rosenfeld "A maximum entropy approach to adaptive statistical language modelling"; Computer Speech and Language, 10:187-228, 1996.

- Mit der Substitution der Merkmalsfunktionen auf der linken Seite in Gleichung 2 ändern sich allerdings auch die Randwerte m_{α} auf deren rechter Seite. Damit wird das ursprüngliche System von Randbedingungen, d.h. die ursprüngliche Trainingsaufgabe bei den
 30

üblichen Ansätzen zum Schätzen der Randwerte verändert; siehe dazu Rosenfeld a.a.O., S. 205, 1.Satz des vorletzten Absatzes.

- Es ist insofern als Nachteil des Standes der Technik festzustellen, dass bei der
- 5 beschleunigten Abwicklung des GIS-Algorithmus die freien Parameter λ auf eine veränderte Trainingsaufgabe trainiert werden. Die auf diese Weise berechneten Parameter λ bewirken bei Einsetzung in Gleichung 3 nur eine unzulängliche Anpassung des Sprachmodells an die ursprüngliche Trainingsaufgabe.
- 10 Ausgehend von diesem Stand der Technik ist es die Aufgabe der Erfindung, ein bekanntes computergestütztes Spracherkennungssystem, eine computergestützte Trainingseinrichtung und ein bekanntes Verfahren zum iterativen Berechnen freier Parameter $\lambda_{\alpha}^{ortho(n)}$ eines Maximum-Entropie-Sprachmodells in dem Spracherkennungssystem derart weiterzubilden, dass sie eine schnelle Berechnung der freien Parameter λ ohne Veränderung der
- 15 ursprünglichen Trainingsaufgabe ermöglichen.

- Diese Aufgabe wird gemäß Patentanspruch 1 dadurch gelöst, dass bei dem bekannten, einleitend beschriebenen Verfahren zur Berechnung der freien Parameter λ nach dem GIS-Algorithmus jeder gewünschte orthogonalisierte Randwert m_{α}^{ortho} durch Linearkombi-
- 20 nation des zugehörigen gewünschten Randwertes m_{α} mit gewünschten Randwerten m_{β} von zu dem Merkmal α höher-reichweitigen Merkmalen β berechnet wird. Hierbei sind m_{α} und m_{β} gewünschte Randwerte der ursprünglichen Trainingsaufgabe.

- Die Verwendung der so berechneten Randwerte m_{α}^{ortho} ermöglicht vorteilhafterweise eine
- 25 verbesserte Approximation der freien Parameter λ und damit eine Verbesserung des Sprachmodells im Hinblick auf die ursprüngliche Trainingsaufgabe. Diese qualitative Verbesserung ist möglich bei gleichzeitiger Realisierung einer hohen Konvergenzgeschwindigkeit für die freien Parameter λ bei deren iterativer Berechnung mit Hilfe des GIS-Algorithmus.

8

Die Verwendung der erfindungsgemäß berechneten gewünschten orthogonalisierten Randwerte m_{α}^{ortho} empfiehlt sich für verschiedene Varianten des GIS-Trainingsalgorithmus, wie sie in den Unteransprüchen 12 und 13 beschrieben sind.

- 5 Die Aufgabe der Erfindung wird weiterhin durch ein Spracherkennungssystem auf Basis des Maximum Entropie Sprachmodells MESM gemäß Patentanspruch 14 und eine Trainingseinrichtung zum Trainieren des MESM gemäß Patentanspruch 15 gelöst.

- 10 Durch Benutzen des erfindungsgemäßen Verfahrens in der Trainingseinrichtung erfolgt eine im Vergleich zum Stand der Technik effektivere Anpassung des MESM in dem Spracherkennungssystem an die individuellen sprachlichen Eigenheiten eines bestimmten Benutzers des Spracherkennungssystems; die Quote mit welcher das Spracherkennungssystem dann die Bedeutungsinhalte in der Sprache des Benutzers richtig erkennt wird wesentlich verbessert.

- 15 Ansonsten entsprechen die Vorteile dieses Spracherkennungssystems und der Trainingseinrichtung den oben für das Verfahren diskutierten Vorteilen.

Der Beschreibung der Erfindung sind folgende Figuren beigelegt, wobei

- 20 Fig. 1a : ein erfindungsgemäßes Verfahren zur Berechnung
 + 1b : eines gewünschten orthogonalisierten Randwertes m_{α}^{ortho} beschreibt;
 Fig. 2a : ein erfindungsgemäßes Verfahren zur Berechnung
 25 + 2b : einer orthogonalisierten Merkmalsfunktion f_{α}^{ortho} beschreibt;
 Fig. 3 : ein Blockschaltbild eines erfindungsgemäßen Spracherkennungssystems;

und

Fig. 4 : einen Merkmalsbaum beschreibt.

Es erfolgt im weiteren zunächst eine detaillierte Beschreibung eines Ausführungsbeispiels der Erfindung unter Bezugnahme auf die Fig. 1a und 1b.

5

Die Fig. 1a und 1b veranschaulichen ein erfindungsgemäßes Verfahren zum Berechnen

eines verbesserten gewünschten orthogonalisierten Randwertes m_{α}^{ortho} für ein Merkmal

$\alpha = \beta_0$ in einem Sprachmodell. In einem ersten Verfahrensschritt S1 werden gemäß

diesem Verfahren alle in dem Sprachmodell definierten Merkmale β_i mit $i = 1 \dots g$

10 bestimmt, welche eine sog. höhere Reichweite aufweisen als das vorbestimmte Merkmal

$\alpha = \beta_0$, d.h. welche dieses an einer vorbestimmten Stelle einschließen. Anschließend wird

in einem Verfahrensschritt S2 für alle Merkmale β_i mit $i = 0 \dots g$, also auch für das

Merkmal $\alpha = \beta_0$, ein gewünschter Randwert m_{β_i} der ursprünglichen Trainingsaufgabe berechnet.

15

Für die Berechnung eines solchen gewünschten Randwertes m_{β_i} sind verschiedene Methoden im Stand der Technik bekannt.

Gemäß einer ersten Methode erfolgt die Berechnung dadurch, dass zunächst eine Häufig-

20 keit $N(\beta_i)$ ermittelt wird, mit welcher die zugehörige binäre Merkmalsfunktion f_{β_i} bei

Anwendung auf einem Trainingskorpus des Sprachmodells den Wert 1 ergibt und dass

anschließend der so ermittelte Häufigkeitswert $N(\beta_i)$ geglättet wird.

Gemäß einer zweiten alternativen Methode erfolgt die Berechnung durch Reduktion von

25 Merkmalsmengen in dem Sprachmodell so lange, bis die Randbedingungen keine Wider-

sprüche mehr aufweisen. Eine derartige Reduktion von Merkmalsmengen muss in der

Praxis sehr umfangreich sein, weil ansonsten das erzeugte Sprachmodell keine Lösung mehr zu der ursprünglichen Trainingsaufgabe darstellt.

Gemäß einer dritten Methode erfolgt die Berechnung durch Benutzung eines sogenannten

30 induzierenden Sprachmodells, wie es in J.Peters und D.Klakow, "Compact Maximum

Entropy Language Models", Proc. ASRU, Keystone, Colorado, 1999 beschrieben ist.

In einem Verfahrensschritt S3 werden nachfolgend alle Merkmale β_i nach ihrer Reichweite sortiert, wobei einem Merkmal β_i mit der größten Reichweite der Index $i = g$ zugeordnet wird. Dabei kann es durchaus vorkommen, dass einzelnen Reichweitenklassen, also z.B. der Klasse Bigramme oder der Klasse Trigramme, mehrere Merkmale β_i zugeordnet werden. In diesen Fällen sind mehrere Merkmale β_i mit unterschiedlichen, aber aufeinanderfolgenden Indizes i ein und derselben Reichweitenklasse zugeordnet, d.h. diese Merkmale haben dann jeweils dieselbe RW und gehören derselben Reichweitenklasse an.

Für den Ablauf des Verfahrens, bei welchem in den nachfolgenden Schritten die einzelnen Merkmale β_i der Reihe nach ausgewertet werden, ist es wichtig, dass die Merkmale nach absteigender (oder gleichbleibender) Reichweite abgearbeitet werden. Im ersten Durchlauf des Verfahrens wird daher mit einem Merkmal β_i begonnen, welches der höchsten Reichweitenklasse zugeordnet ist; vorzugsweise wird $i = g$ gesetzt (siehe Verfahrensschritte S4 und S5 in Fig. 1a).

In einem nachfolgenden Verfahrensschritt S6 wird geprüft, ob es zu dem aktuell ausgewählten Merkmal β_i höher- reichweitige Merkmale β_k mit $i < k \leq g$ gibt, welche das Merkmal β_i einschließen. Beim ersten Durchlauf gehört das Merkmal β_i mit $i=g$, wie oben gesagt, automatisch der Klasse mit der höchsten Reichweite an und deshalb ist die Abfrage in Verfahrensschritt S6 für dieses Merkmal β_i zu verneinen. In diesem Fall springt das Verfahren zu Verfahrensschritt S8, wo ein Parameter X zu Null gesetzt wird. Es erfolgt daraufhin eine Berechnung eines verbesserten gewünschten orthogonalisierten Randwertes m_{β}^{ortho} für das Merkmal β_i (beim ersten Durchlauf mit $i=g$) gemäß Verfahrensschritt S9. Wie dort ersichtlich, wird dieser Randwert für das Merkmal β_i dem in Schritt S2 berechneten gewünschten Randwert m_{β_i} gleichgesetzt, wenn der Parameter $X=0$ ist (dies ist z.B. beim ersten Durchlauf der Fall).

Die Verfahrensschritte S5 bis S11 werden daraufhin sukzessive für alle Merkmale β_{i-1} mit $i-1 = g-1 \dots 0$ wiederholt. In Verfahrensschritt S10 findet eine dafür notwendige Neuinitialisierung des Indexes i statt und in Verfahrensschritt S11 erfolgt eine Abfrage, ob alle

Merkmale β_i mit $i = 0 \dots g$ abgearbeitet worden sind.

Für alle Merkmale β_i , für welche höher-reichweitige Merkmale β_k mit $i < k \leq g$ existieren, ist die Abfrage in Verfahrensschritt S6 mit "Ja" zu beantworten. Der Parameter X wird
 5 dann nicht zu Null gesetzt, sondern berechnet sich gemäß Verfahrensschritt S7 durch Aufsummierung der entsprechenden, in vorherigen Durchläufen jeweils in Verfahrensschritt S9 berechneten verbesserten gewünschten orthogonalisierten Randwerten $m_{\beta k}^{ortho}$ für die jeweils höher-reichweitigen Merkmale β_k .

10 Sobald in Verfahrensschritt S11 festgestellt worden ist, dass der gewünschte orthogonalisierte Randwert $m_{\beta 0}^{ortho}$ in Verfahrensschritt S9 berechnet worden ist, wird dieser in Verfahrensschritt S12 als m_{α}^{ortho} ausgegeben.

Das soeben ausführlich beschriebene erfindungsgemäße Verfahren zur Berechnung des
 15 verbesserten gewünschten orthogonalisierten Randwertes m_{α}^{ortho} lässt sich in Kurzschreibweise gemäß folgender Formel zusammenfassen:

$$m_{\alpha}^{ortho} = m_{\alpha} - \sum_{(*)} m_{\beta}^{ortho} \quad (4)$$

20 Die Summe (*) umfasst alle höher-reichweitigen Merkmale β , welche das vorbestimmte Merkmal α einschließen. Zur Berechnung des Randwertes m_{β}^{ortho} ist die genannte Formel quasi rekursiv für jedes Merkmal β immer wieder anzuwenden, bis für bestimmte Merkmale, nämlich jene mit der höchsten Reichweite, der Summenterm verschwindet, weil zu diesen Merkmalen keine höher-reichweitigen existieren. Die gewünschten orthogonalisierten Randwerte für die höchstreichweitigen Merkmale β_k entsprechen dann jeweils den
 25 ursprünglichen gewünschten Randwerten $m_{\beta k}$.

Die Anwendung des erfindungsgemäßen Verfahrens gemäß den Fig. 1a und 1b soll nachfolgend durch Anwendung auf dem folgenden beispielhaften Trainingskorpus eines

12

Sprachmodells erläutert werden. Der Trainingskorpus lautet:

" THAT WAS A RED
 OR A GREEN HOUSE
 5 OR A BLUE HOUSE
 THIS IS A WHITE HOUSE AND
 THAT IS THE WHITE HOUSE"

Das Trainingskorpus besteht aus $N = 23$ einzelnen Worten. Es sei vorausgesetzt, dass in
 10 dem Sprachmodell die gewünschten Unigramm-, Bigramm- und Trigramm-Merkmale
 gemäß Fig. 4 vordefiniert sind.

Es kann dann durch Anwendung der normalen Merkmalsfunktion f_α auf das Trainings-
 korpus festgestellt werden, dass die Unigramme, Bigramme und Trigramme gemäß Fig. 4
 15 mit folgenden Häufigkeiten in dem Trainingskorpus vorhanden sind:

Unigramme:

	A	4
20	HOUSE	4
	IS	2
	OR	2
	THAT	2
	WHITE	2

25

Bigramme:

	A	WHITE	1
	OR	A	2
	WHITE	HOUSE	2

30

Trigramme:

	A	WHITE	HOUSE	1
--	---	-------	-------	---

13

In dem hier gezeigten Beispiel soll der verbesserte gewünschte orthogonalisierte Randwert

m_{α}^{ortho} für das Merkmal $\alpha = \text{"HOUSE"}$ berechnet werden. Dafür sind zunächst gemäß

Verfahrensschritt S1 in Fig. 1a alle zu dem Merkmal α höher-reichweitigen Merkmale zu bestimmen. Es sind dies gemäß Fig. 4 das Bigramm "WHITE HOUSE" und das Tri-

5 gramm "A WHITE HOUSE". Gemäß Verfahrensschritt S2 sind nun für diese höher-

reichweitigen Merkmale, aber auch für das Merkmal α die normalen gewünschten Randwerte zu berechnen, z.B. indem die jeweiligen, oben festgestellten Häufigkeiten geglättet werden. Diese Glättung erfolgt hier beispielhaft durch Subtraktion des Wertes 0,1. Damit ergeben sich folgende normale gewünschte Randwerte:

10

$$m_{\alpha} : \text{"HOUSE"} = 4 - 0,1 = 3,9$$

$$m_{\beta 1} : \text{"WHITE HOUSE"} = 2 - 0,1 = 1,9$$

15 $m_{\beta 2} : \text{"A WHITE HOUSE"} = 1 - 0,1 = 0,9.$

Die Merkmale α , β_1 , β_2 werden nun ihrer Reichweite nach sortiert und es werden - beginnend mit dem längst- reichweitigen Merkmal - die jeweiligen verbesserten gewünschten orthogonalisierten Randwerte gemäß Formel (6) bzw. gemäß

20 Verfahrensschritt S7 - S9 in Fig. 1a und 1b berechnet:

$$m_{\beta 2}^{ortho} = m_{\beta 2} = 0,9 \quad (5)$$

$$m_{\beta 1}^{ortho} = m_{\beta 1} - m_{\beta 2}^{ortho} = 1,9 - 0,9 = 1 \quad (6)$$

25

Schließlich berechnet sich der verbesserte gewünschte orthogonalisierte Randwert m_{α}^{ortho} für das Merkmal α zu:

$$m_{\alpha}^{ortho} = m_{\alpha} - m_{\beta 1}^{ortho} - m_{\beta 2}^{ortho} = 3,9 - 1 - 0,9 = 2 \quad (7)$$

Der so erfindungsgemäß berechnete orthogonalisierte Randwert m_{α}^{ortho} ermöglicht eine ausreichend genaue Berechnung der freien Parameter λ und damit der Wahrscheinlichkeit gemäß Formel (1) im Hinblick auf eine ursprüngliche Trainingsaufgabe bei gleichzeitig großer Rechengeschwindigkeit bei Verwendung in dem GIS-Trainingsalgorithmus.

5

Nachfolgend wird die Verwendung des erfindungsgemäß berechneten Randwertes m_{α}^{ortho} für drei unterschiedliche Varianten des GIS-Trainingsalgorithmus dargestellt.

Bei einer ersten Variante des GIS-Trainingsalgorithmus hat die mathematische Funktion G gemäß Gleichung 1 bei Verwendung des erfindungsgemäß berechneten orthogonalisierten Randwertes m_{α}^{ortho} folgende Gestalt:

$$\begin{aligned} \lambda_{\alpha}^{ortho(n+1)} &= G(\lambda_{\alpha}^{ortho(n)}, m_{\alpha}^{ortho}, \dots) \\ &= \lambda_{\alpha}^{ortho(n)} + t_{\alpha}^{ortho} \cdot \log \left(\frac{[t_{\alpha}^{ortho} \cdot m_{\alpha}^{ortho} + b_{\alpha}]}{[t_{\alpha}^{ortho} \cdot m_{\alpha}^{ortho(n)} + b_{\alpha}]} \cdot \frac{1 - \sum_{\gamma} [t_{\gamma}^{ortho} \cdot m_{\gamma}^{ortho(n)} + b_{\gamma}]}{1 - \sum_{\gamma} [t_{\gamma}^{ortho} \cdot m_{\gamma}^{ortho} + b_{\gamma}]} \right) \end{aligned} \quad (8)$$

15 wobei:

n : einen Iterationsparameter;
 α : ein gerade betrachtetes Merkmal;
 γ : alle Merkmale in dem Sprachmodell;
 $t_{\alpha}^{ortho}, t_{\gamma}^{ortho}$: die Größe eines Konvergenzschrittes;
 $m_{\alpha}^{ortho}, m_{\gamma}^{ortho}$: gewünschte orthogonalisierte Randwerte in dem MESM für die Merkmale α und γ ;

$m_{\alpha}^{ortho(n)}, m_{\gamma}^{ortho(n)}$: iterative Näherungswerte für die gewünschten Randwerte
 $m_{\alpha}^{ortho}, m_{\gamma}^{ortho}$; und

25

75

$b\alpha$ und $b\gamma$: Konstanten

bezeichnen.

- 5 Die Berechnung der Konvergenzschrittgrößen t und der iterativen Näherungswerte für die gewünschten Randwerte m erfolgt - wie nachfolgend gezeigt wird - durch Verwendung einer erfindungsgemäß definierten orthogonalisierten Merkmalsfunktion f_a^{ortho} , welche wie folgt lautet:

$$10 \quad f_a^{ortho} = f\alpha - \sum_{(*)} f_\beta^{ortho} \quad (9)$$

- Es sei an dieser Stelle betont, dass die gemäß Gleichung 9 erfindungsgemäß berechnete orthogonale Merkmalsfunktion f_a^{ortho} in ihrem Betrag der von Rosenfeld a.a.O. vorgeschlagenen Merkmalsfunktion entspricht. Allerdings erfolgt ihre erfindungsgemäße
- 15 Berechnung, wie in den Fig. 2a und 2b veranschaulicht, gänzlich anders. Das Berechnungsverfahren läuft analog ab zu dem in den Fig. 1a und 1b beschriebenen Verfahren zur Berechnung der gewünschten orthogonalisierten Randwerte m_a^{ortho} , wobei lediglich das Symbol für den Randwert m durch das Symbol für die Merkmalsfunktion f und der Parameter X durch die Funktion F zu ersetzen ist. Um Wiederholungen zu vermeiden,
- 20 wird an dieser Stelle für Erläuterungen des Verfahrens gemäß Fig. 2a und 2b auf die Beschreibung der Fig. 1a und 1b verwiesen.

- Mit der so erfindungsgemäß berechneten orthogonalisierten Merkmalsfunktion f_a^{ortho} bzw. f_β^{ortho} berechnet sich die Größe der Konvergenzschröße t_a^{ortho} und t_γ^{ortho} in Gleichung
- 25 8 wie folgt:

$$t_\alpha^{ortho} = t_\gamma^{ortho} = 1/M^{ortho} \quad \text{mit} \quad M^{ortho} = \max_{(h,w)} \left(\sum_\beta f_\beta^{ortho}(h,w) \right) \quad (10)$$

16

wobei M_{ortho} für binäre Merkmalsfunktionen f_{β}^{ortho} die maximale Anzahl von Funktionen repräsentiert, welche für dasselbe Argument (h, w) den Wert 1 liefern.

- Weiterhin berechnet sich mit der erfindungsgemäß definierten Merkmalsfunktion f_{α}^{ortho}
- 5 der iterative Näherungswert $m_{\alpha}^{ortho(n)}$ für den gewünschten orthogonalisierten Randwert m_{α}^{ortho} bei analoger Anwendung der Gleichung (2) wie folgt:

$$m_{\alpha}^{ortho(n)} = \sum_{(h, w)} N(h) \cdot p^{(n)}(w | h) \cdot f_{\alpha}^{ortho}(h, w) ; \quad (11)$$

- 10 wobei:

$N(h)$: die Häufigkeit der Historie h im Trainingskorpus; und

$p^{(n)}(w|h)$: einen Iterationswert für die Wahrscheinlichkeit $p(w|h)$, mit welcher sich ein vorgegebenes Wort w an eine bisherige Wortfolge h (Historie) anschließt;

- 15

bezeichnet. Hierbei benutzt $p^{(n)}(w|h)$ die Parameterwerte $\lambda_{\alpha}^{ortho(n)}$.

- Die Verwendung des erfindungsgemäß berechneten verbesserten gewünschten orthogonalisierten Randwertes m_{α}^{ortho} empfiehlt sich weiterhin für eine zweite Variante des GIS-
- 20 Trainingsalgorithmus. Hierbei sind die Merkmale des MESM in m Gruppen A_i aufgeteilt, und pro Iteration werden nur die Parameter λ_{α}^{ortho} der Merkmale α , aus einer der Gruppen gemäß der folgenden Formel verändert:

$$\begin{aligned} \lambda_{\alpha}^{ortho(n+1)} &= G(\lambda_{\alpha}^{ortho(n)}, m_{\alpha}^{ortho}, \dots) \\ &= \lambda_{\alpha}^{ortho(n)} + t_{\alpha}^{ortho} \cdot \log \left(\frac{m_{\alpha}^{ortho}}{m_{\alpha}^{ortho(n)}} \cdot \frac{1 - \sum_{\beta \in A_i(n)} (t_{\beta} \cdot m_{\beta}^{ortho(n)})}{1 - \sum_{\beta \in A_i(n)} (t_{\beta} \cdot m_{\beta}^{ortho})} \right) \end{aligned} \quad (12)$$

77

wobei:

n : der Iterationsparameter

$A_i(n)$: eine in dem n 'ten Iterationsschritt ausgewählte

5 Merkmalsgruppe $A_i(n)$ mit $1 \leq i \leq m$;

α : ein gerade betrachtetes Merkmal aus der gerade
ausgewählten Merkmalsgruppe $A_i(n)$;

β : alle Merkmale aus der Merkmalsgruppe $A_i(n)$;

t_α^{ortho} , t_β^{ortho} : die Größe eines Konvergenzschrittes mit

10
$$t_\alpha^{ortho} = t_\beta^{ortho} = 1/M_{i(n)}^{ortho} \text{ mit}$$

$$M_{i(n)}^{ortho} = \max_{(h,w)} \left(\sum_{\beta \in A_i(n)} f_\beta^{ortho}(h,w) \right)$$

wobei

$M_{i(n)}^{ortho}$ für binäre Funktionen f_β^{ortho} die maximale Anzahl von
Funktionen aus der Merkmalsgruppe $A_i(n)$ repräsentiert, welche für
15 dasselbe Argument (h,w) den Wert 1 liefern;

m_α^{ortho} , m_β^{ortho} : gewünschte orthogonalisierte Randwerte in dem MESM für
die Merkmale α bzw. β ;

$m_\alpha^{ortho(n)}$, $m_\beta^{ortho(n)}$: iterative Näherungswerte für die gewünschten Randwerte

$$m_\alpha^{ortho}, m_\beta^{ortho};$$

20

Die Gruppe $A_i(n)$ von Merkmalen α , deren Parameter λ_α^{ortho} im aktuellen Iterationsschritt
angepasst werden, durchläuft dabei zyklisch gemäß $i(n)=n(\text{mod } m)$ alle m Gruppen.

Die Verwendung des erfindungsgemäß berechneten gewünschten orthogonalisierten

25 Randwertes m_α^{ortho} empfiehlt sich weiterhin für eine dritte Variante des GIS-Trainings-
algorithmus, welche sich von der zweiten Variante lediglich dadurch unterscheidet, dass die
bei jedem Iterationsschritt zu verwendende Merkmalsgruppe $A_i(n)$ nicht zyklisch, sondern

78

nach einem vorbestimmten Kriterium $D_i^{(n)}$ ausgewählt wird.

In Fig. 3 ist schließlich ein Spracherkennungssystem 10 gemäß der vorliegenden Erfindung dargestellt, welches auf Basis des sog. Maximum-Entropie-Sprachmodells arbeitet. Es
5 umfasst eine Erkennungseinrichtung 12, welche den Bedeutungsinhalt von bereitgestellten Sprachsignalen zu erkennen versucht. Die Sprachsignale werden dem Spracherkennungssystem üblicherweise als Ausgangssignale eines Mikrophons 20 bereitgestellt. Die Erkennungseinrichtung 12 erkennt den Bedeutungsinhalt der Sprachsignale dadurch, dass sie
10 Muster in dem empfangenen akustischen Signal auf vordefinierte Erkennungssymbole, wie z.B. bestimmte Worte, Handlungen oder Ereignisse, mit Hilfe des implementierten Maximum Entropie Sprachmodells MESM abbildet. Schließlich gibt die Erkennungseinrichtung 12 ein Ausgangssignal aus, welches den in dem Sprachsignal erkannten Bedeutungsinhalt repräsentiert und zur Ansteuerung von verschiedensten Geräten, z.B. eines Textverarbeitungsprogramms oder eines Telefons nach Maßgabe durch diesen Bedeutungsinhalt dienen kann.
15

Für eine möglichst fehlerfreie Ansteuerung der Geräte im Sinne des Bedeutungsinhaltes von steuernder Sprachinformation ist es erforderlich, dass das Spracherkennungssystem 10 die Bedeutungsinhalte der auszuwertende Sprache mit einer möglichst hohen Quote richtig
20 erkennt. Dazu ist eine möglichst gute Anpassung des Sprachmodells an die sprachlichen Besonderheiten des Sprechers, d.h. des Benutzers des Spracherkennungssystems erforderlich. Diese Anpassungsaufgabe leistet eine Trainingseinrichtung 14, welche entweder extern oder in das Spracherkennungssystem 10 integriert betrieben werden kann. Genauer gesagt dient die Trainingseinrichtung 14 zum Anpassen des MESM in dem Spracherkennungssystem 10 an wiederkehrende statistische Muster in der Sprache eines bestimmten Benutzers.
25

Sowohl die Erkennungseinrichtung 12 wie auch die Trainingseinrichtung 14 sind üblicherweise, aber nicht notwendigerweise, als Softwaremodule ausgebildet und laufen auf
30 einem geeigneten Computer (nicht gezeigt) ab.

7p

PATENTANSPRÜCHE

1. Verfahren zum iterativen Berechnen freier Parameter $\lambda_{\alpha}^{ortho(n)}$ eines Maximum-Entropie-Sprachmodells MESM mit Hilfe des Generalised-Iterative-Scaling-Trainingsalgorithmus gemäß folgender Formel:

$$5 \quad \lambda_{\alpha}^{ortho(n+1)} = G(\lambda_{\alpha}^{ortho(n)}, m_{\alpha}^{ortho}, \dots)$$

wobei:

n : einen Iterationsparameter;

G : eine mathematische Funktion;

10 α : ein Merkmal in dem MESM; und

m_{α}^{ortho} : einen gewünschten orthogonalisierten Randwert in dem MESM für das Merkmal α ;

dadurch gekennzeichnet,

dass der gewünschte orthogonalisierte Randwert m_{α}^{ortho} durch Linearkombination des

15 gewünschten Randwertes m_{α} mit gewünschten Randwerten m_{β} von zu dem Merkmal α höher-reichweitigen Merkmalen β berechnet wird.

2. Verfahren nach Anspruch 1,

dadurch gekennzeichnet,

20 dass die Berechnung des gewünschten orthogonalisierten Randwertes m_{α}^{ortho} für das Merkmal $\alpha = \beta_0$ folgende Schritte umfasst:

- a) Auswählen aller Merkmale β_i mit $i=1\dots g$ in dem Sprachmodell, welche eine höhere Reichweite RW aufweisen als das Merkmal $\alpha=\beta_0$ und dieses jeweils einschließen;
- b) Berechnen von gewünschten Randwerten m_{β_i} für die Merkmale β_i mit $i=0\dots g$;
- 5 c) Sortieren der Merkmale β_i mit $i=0\dots g$ nach ihrer RW;
- d) Auswahl eines der Merkmale β_i mit der höchsten RW;
- e) Prüfen, ob es zu dem ausgewählten Merkmal β_i andere Merkmale β_k höherer RW gibt,
10 welche das Merkmal β_i einschließen;
- f1) Wenn ja, dann Definieren eines Parameters X als Linearkombination der beim letzten Durchlauf der Schritte e) bis g) in Schritt g) berechneten orthogonalisierten Randwerte $m_{\beta_k}^{ortho}$ für alle im zuletzt ausgeführten Schritt e) bestimmten höher-reichweitigen
15 Merkmale β_k ;
- f2) Wenn nein, dann Definieren des Parameters X zu $X = 0$;
- g) Berechnen des gewünschten orthogonalisierten Randwertes $m_{\beta_i}^{ortho}$ für das Merkmal β_i
20 durch arithmetisches Verknüpfen des gewünschten Randwertes m_{β_i} mit dem Parameter X;
und
- h) Wiederholen der Schritte e) bis g) für das Merkmal β_{i-1} , dessen RW kleiner oder gleich der RW des Merkmals β_i ist, solange bis der
25 gewünschte orthogonalisierte Randwert $m_{\beta_0}^{ortho} = m_{\alpha}^{ortho}$ mit $i=0$ in Schritt g) berechnet worden ist.

3. Verfahren nach Anspruch 2,

dadurch gekennzeichnet,

dass die Berechnung des Parameters X in Schritt f1) gemäß folgender Formel erfolgt:

$$X = \sum_k m_{\beta k}^{ortho}$$

5

4. Verfahren nach Anspruch 3,

dadurch gekennzeichnet,

dass die Berechnung des gewünschten orthogonalisierten Randwertes $m_{\beta i}^{ortho}$ in Schritt g)

10 gemäß der folgenden Formel erfolgt:

$$m_{\beta i}^{ortho} = m_{\beta i} - X$$

15 5. Verfahren nach Anspruch 2,

dadurch gekennzeichnet,

dass die Berechnung von gewünschten Randwerten $m_{\beta i}$ für die Merkmale β_i mit $i = 0, \dots, g$ in Schritt b) durch jeweiliges Berechnen der Häufigkeit $N(\beta_i)$, mit welcher das Merkmal β_i in einem Trainingskorpus enthalten ist, und durch nachfolgendes Glätten des

20 berechneten Häufigkeitswertes $N(\beta_i)$ erfolgt.

25

22

6. Verfahren nach Anspruch 5,

dadurch gekennzeichnet,

dass die Berechnung der Häufigkeit $N(\beta_i)$ durch Anwenden einer binären Merkmalsfunktion f_{β_i} auf den Trainingskorpus erfolgt, wobei f_{β_i} definiert ist als:

5

$$f_{\beta_i}(h, w) = \begin{cases} 1 & \text{falls } \beta_i \text{ zur Wortfolge } (h, w) \text{ passt} \\ 0 & \text{sonst} \end{cases}$$

und wobei $f_{\beta_i}(h, w)$ eine Aussage darüber macht, ob das Merkmal β_i ein durch die Wortfolge (h, w) vorgegebenes Muster richtig beschreibt.

10

7. Verfahren nach Anspruch 1,

dadurch gekennzeichnet,

dass die mathematische Funktion G als weitere Variable die Größe eines

Konvergenzschrittes t_{α}^{ortho} mit:

15

$$t_{\alpha}^{ortho} = 1/M^{ortho}$$

aufweist, wobei

M^{ortho} : für binäre Funktionen f_{α}^{ortho} die maximale Anzahl von Funktionen repräsentiert, welche für dasselbe Argument (h, w) den Wert 1 liefern.

20

8. Verfahren nach Anspruch 7,

dadurch gekennzeichnet,

dass die Merkmalsfunktion f_{α}^{ortho} durch Linearkombination einer Merkmalsfunktion f_{α}

mit orthogonalisierten Merkmalsfunktionen f_{β}^{ortho} von zu dem Merkmal α höher-

25 reichweitigen Merkmalen β berechnet wird.

9. Verfahren nach Anspruch 8,

dadurch gekennzeichnet,

dass die Berechnung der orthogonalisierten Merkmalsfunktion f_{α}^{ortho} für das Merkmal $\alpha = \beta_0$ folgende Schritte umfasst:

5

a) Auswählen aller Merkmale β_i mit $i=1\dots g$ in dem Sprachmodell, welche eine höhere Reichweite RW aufweisen als das Merkmal $\alpha = \beta_0$ und dieses jeweils einschließen;

b) Berechnen von Merkmalsfunktionen f_{β_i} für die Merkmale β_i mit $i=0\dots g$;

10

c) Sortieren der Merkmale β_i mit $i=0\dots g$ nach ihrer RW;

d) Auswahl eines der Merkmale β_i mit der höchsten RW;

15 e) Prüfen, ob es zu dem ausgewählten Merkmal β_i andere Merkmale β_k höherer RW gibt, welche das Merkmal β_i einschließen;

f1) Wenn ja, dann Definieren einer Funktion F als Linearkombination der beim letzten Durchlauf der Schritte e) bis g) in Schritt g) berechneten orthogonalisierten

20 Merkmalsfunktion $f_{\beta_k}^{ortho}$ für alle im zuletzt ausgeführten Schritt e) bestimmten höherreichweitigen Merkmale β_k ;

f2) Wenn nein, dann Definieren der Funktion F zu $F = 0$;

25 g) Berechnen der orthogonalisierten Merkmalsfunktion $f_{\beta_i}^{ortho}$ für das Merkmal β_i durch arithmetisches Verknüpfen der Merkmalsfunktion f_{β_i} mit der Funktion F; und

h) Wiederholen der Schritte e) bis g) für das Merkmal β_{i-1} , dessen Reichweite kleiner oder

24

gleich der Reichweite des Merkmals β_i ist, solange bis die orthogonalisierte

Merkmalsfunktion $f_{\beta 0}^{ortho} = f_{\alpha}^{ortho}$ mit $i=0$ in Schritt g) berechnet worden ist.

10. Verfahren nach Anspruch 9,

5 dadurch gekennzeichnet,

dass die Berechnung der Funktion F in Schritt f1) gemäß folgender Formel erfolgt:

$$F = \sum_k f_{\beta k}^{ortho}$$

10 11. Verfahren nach Anspruch 9,

dadurch gekennzeichnet,

dass die Berechnung der orthogonalisierten Merkmalsfunktion $f_{\beta i}^{ortho}$ in Schritt g) gemäß der folgenden Formel erfolgt:

15
$$f_{\beta i}^{ortho} = f_{\beta i} - F$$

12. Verfahren nach Anspruch 1,

dadurch gekennzeichnet,

dass die mathematische Funktion G folgende Gestalt hat:

20

$$\begin{aligned} \lambda_{\alpha}^{ortho(n+1)} &= G(\lambda_{\alpha}^{ortho(n)}, m_{\alpha}^{ortho}, \dots) \\ &= \lambda_{\alpha}^{ortho(n)} + t_{\alpha}^{ortho} \cdot \log \left(\frac{[t_{\alpha}^{ortho} \cdot m_{\alpha}^{ortho} + b_{\alpha}]}{[t_{\alpha}^{ortho} \cdot m_{\alpha}^{ortho(n)} + b_{\alpha}]} \cdot \frac{1 - \sum_{\gamma} [t_{\gamma}^{ortho} \cdot m_{\gamma}^{ortho(n)} + b_{\gamma}]}{1 - \sum_{\gamma} [t_{\gamma}^{ortho} \cdot m_{\gamma}^{ortho} + b_{\gamma}]} \right) \end{aligned}$$

wobei:

α : ein gerade betrachtetes Merkmal;

25 γ : alle Merkmale in dem Sprachmodell;

25

 $t_{\alpha}^{ortho}, t_{\gamma}^{ortho}$: die Größe eines Konvergenzschrilles mit $t_{\alpha}^{ortho} \approx t_{\gamma}^{ortho} = 1/M^{ortho}$ mit

$$M^{ortho} = \max_{(h,w)} \left(\sum_{\beta} f_{\beta}^{ortho}(h,w) \right);$$

wobei M^{ortho} für binäre Funktionen f_{β}^{ortho} die maximale Anzahl von Funktionen repräsentiert, welche für dasselbe Argument (h,w) den Wert 1 liefern;

5.

 $m_{\alpha}^{ortho}, m_{\gamma}^{ortho}$: gewünschte orthogonalisierte Randwerte in dem MESM für die Merkmale α und γ ; $m_{\alpha}^{ortho(n)}, m_{\gamma}^{ortho(n)}$

: iterative Näherungswerte für die gewünschten Randwerte

 $m_{\alpha}^{ortho}, m_{\gamma}^{ortho}$; und10 b_{α} und b_{γ} : Konstanten

bezeichnen.

13. Verfahren nach Anspruch 1,

15 dadurch gekennzeichnet,

dass die mathematische Funktion folgende Gestalt hat:

$$\begin{aligned} \lambda_{\alpha}^{ortho(n+1)} &= G(\lambda_{\alpha}^{ortho(n)}, m_{\alpha}^{ortho}, \dots) \\ &= \lambda_{\alpha}^{ortho(n)} + t_{\alpha}^{ortho} \cdot \log \left(\frac{m_{\alpha}^{ortho}}{m_{\alpha}^{ortho(n)}} \cdot \frac{1 - \sum_{\beta \in A_i(n)} (t_{\beta} \cdot m_{\beta}^{ortho(n)})}{1 - \sum_{\beta \in A_i(n)} (t_{\beta} \cdot m_{\beta}^{ortho})} \right) \end{aligned}$$

20 wobei:

 n : der Iterationsparameter $A_i(n)$: eine in dem n 'ten Iterationsschritt ausgewählte Merkmalsgruppe $A_i(n)$ mit $1 \leq i \leq m$;

α : ein gerade betrachtetes Merkmal aus der gerade ausgewählten Merkmalsgruppe $A_i(n)$;

β : alle Merkmale aus der Merkmalsgruppe $A_i(n)$;

$t_\alpha^{ortho}, t_\beta^{ortho}$: die Größe eines Konvergenzschrittes mit

$$t_\alpha^{ortho} = t_\beta^{ortho} = 1/M_{i(n)}^{ortho} \text{ mit}$$

$$M_{i(n)}^{ortho} = \max_{(h,w)} \left(\sum_{\beta \in A_i(n)} f_\beta^{ortho}(h,w) \right)$$

5

wobei $M_{i(n)}^{ortho}$ für binäre Funktionen f_β^{ortho} die maximale Anzahl von Funktionen aus der Merkmalsgruppe $A_i(n)$ repräsentiert, welche für dasselbe Argument (h,w) den Wert 1 liefern;

$m_\alpha^{ortho}, m_\beta^{ortho}$: gewünschte orthogonalisierte Randwerte in dem MESM für die Merkmale α bzw. β ;

10

$m_\alpha^{ortho(n)}, m_\beta^{ortho(n)}$: iterative Näherungswerte für die gewünschten Randwerte

$$m_\alpha^{ortho}, m_\beta^{ortho};$$

wobei die Auswahl der Gruppe $A_i(n)$ von Merkmalen α , deren zugehörige Parameter

15 λ_α^{ortho} in einem aktuellen Iterationsschritt angepasst werden, entweder zyklisch oder nach einem vorgegebenen Kriterium erfolgt.

14. Spracherkennungssystem (10) mit:

- 20 einer Erkennungseinrichtung (12) zum Erkennen des Bedeutungsinhaltes eines von einem Mikrophon (20) aufgenommenen und bereitgestellten akustischen Signals, insbesondere eines Sprachsignals, durch Abbilden von Teilen dieses Signals auf vordefinierte Erkennungssymbole, wie sie von dem implementierten Maximum Entropie Sprachmodell MESM angeboten werden, und zum Erzeugen von Ausgangssignalen, welche den erkannten Bedeutungsinhalt repräsentieren; und
- 25 einer Trainingseinrichtung (14) zum Anpassen des MESM an wiederkehrende statistische Muster in der Sprache eines bestimmten Benutzers des Spracherkennungssystems (10);

dadurch gekennzeichnet,

dass die Trainingseinrichtung (14) freie Parameter λ in dem MESM gemäß dem Verfahren nach Anspruch 1 berechnet.

- 5 15. Trainingseinrichtung (14) zum Anpassen des Maximum Entropie Sprachmodells MESM in einem Spracherkennungssystem (10) an wiederkehrende statistische Muster in der Sprache eines bestimmten Benutzers des Spracherkennungssystems (10),

dadurch gekennzeichnet,

- 10 dass die Trainingseinrichtung (14) freie Parameter λ in dem MESM gemäß dem Verfahren nach Anspruch 1 berechnet.

ZUSAMMENFASSUNG

Spracherkennungssystem, Trainingseinrichtung und Verfahren zum iterativen Berechnen freier Parameter eines Maximum-Entropie-Sprachmodells

Die Erfindung betrifft ein Spracherkennungssystem und ein Verfahren zum iterativen

- 5 Berechnen freier Parameter $\lambda_{\alpha}^{ortho(n)}$ eines Maximum-Entropie-Sprachmodells MESM mit Hilfe des Generalised-Iterative-Scaling-Trainingsalgorithmus in einem computergestützten Spracherkennungssystem gemäß der Formel $\lambda_{\alpha}^{ortho(n+1)} = G(\lambda_{\alpha}^{ortho(n)}, m_{\alpha}^{ortho}, \dots)$, wobei n einen Iterationsparameter, G eine mathematische Funktion, α ein Merkmal in dem MESM und m_{α}^{ortho} einen gewünschten orthogonalisierten Randwert in dem MESM für das
- 10 Merkmal α bezeichnet. Es ist die Aufgabe der Erfindung, das System und das Verfahren derart weiterzubilden, dass sie eine schnelle Berechnung des freien Parameters λ ohne Veränderung der ursprünglichen Trainingsaufgabe ermöglichen. Erfindungsgemäß wird diese Aufgabe dadurch gelöst, dass der gewünschte orthogonalisierte Randwert m_{α}^{ortho} durch Linearkombination des gewünschten Randwertes m_{α} mit gewünschten Randwerten
- 15 m_{β} von zu dem Merkmal α höher-reichweitigen Merkmalen β berechnet wird. Hierbei sind m_{α} und m_{β} gewünschte Randwerte der ursprünglichen Trainingsaufgabe.

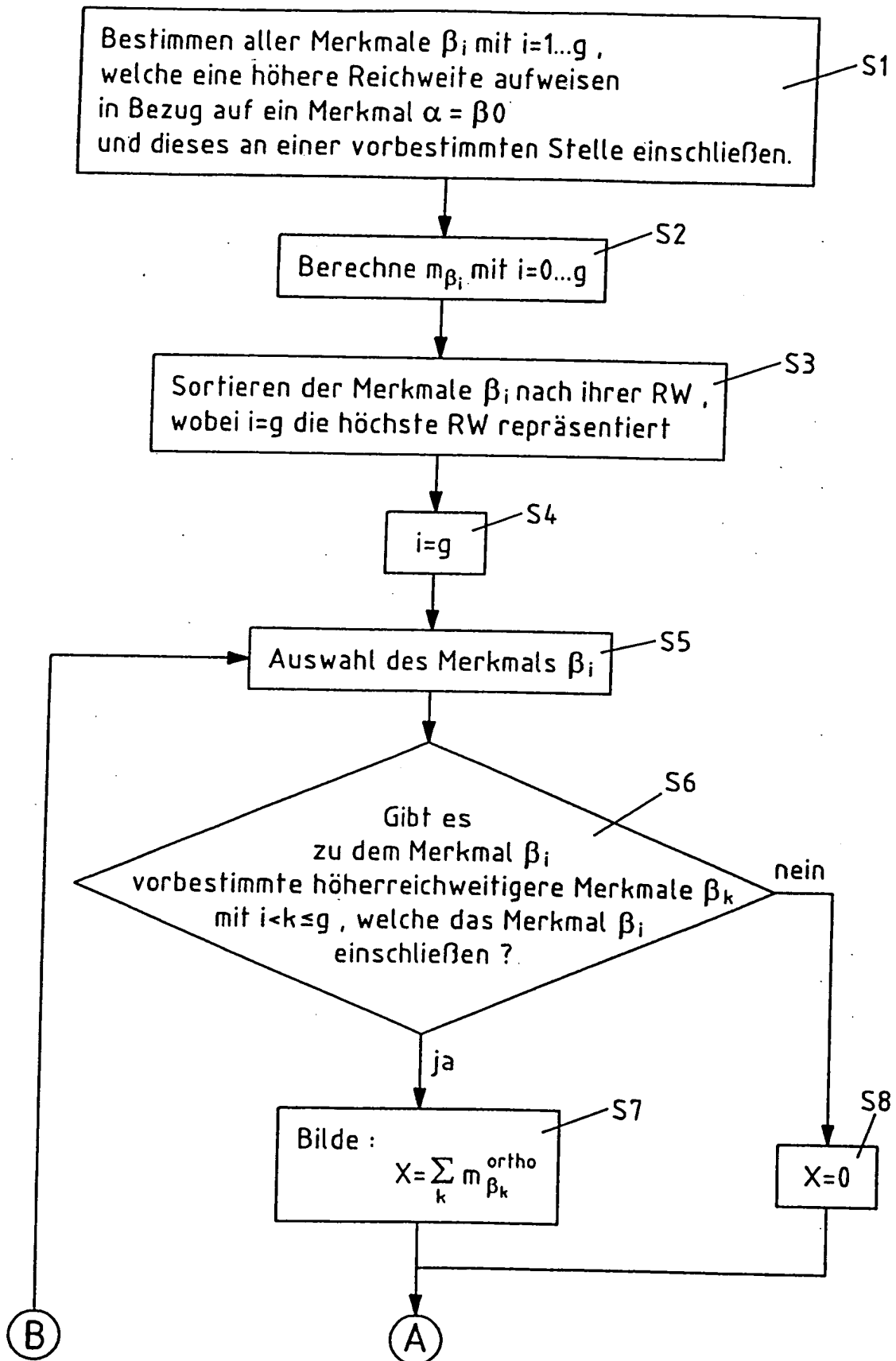


Fig.1a

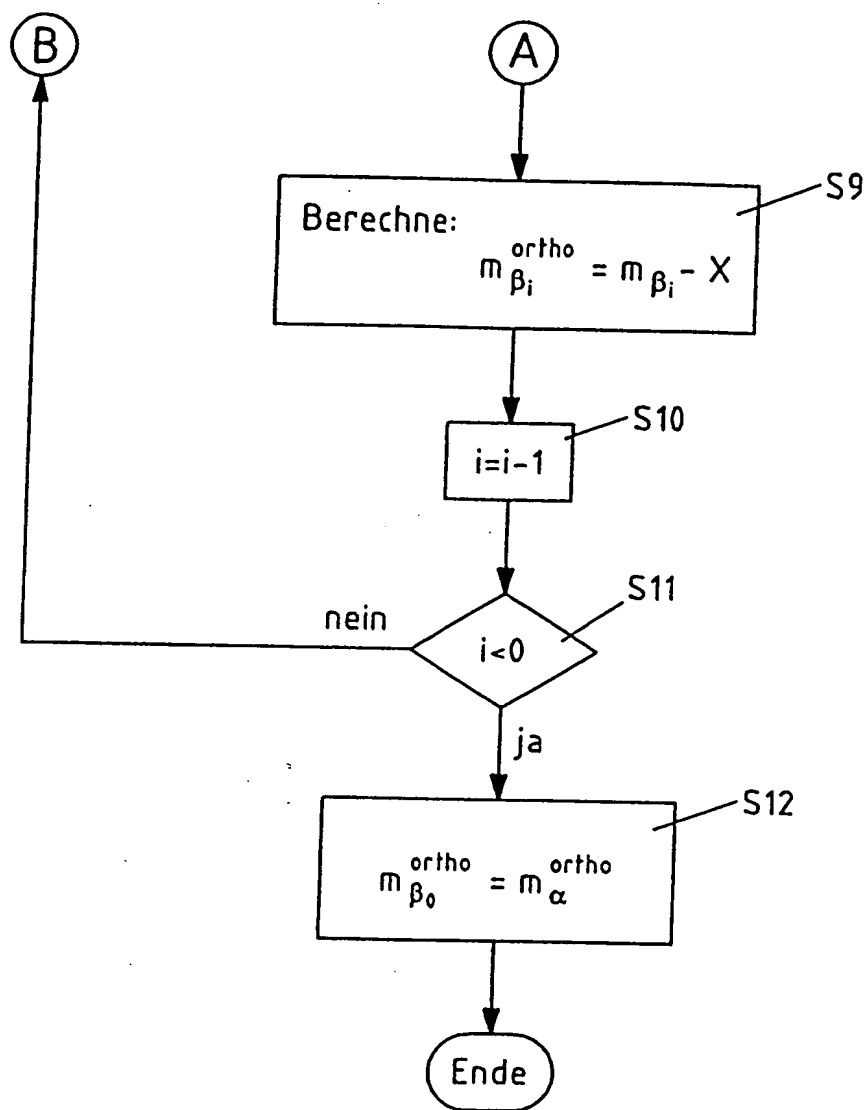


Fig.1b

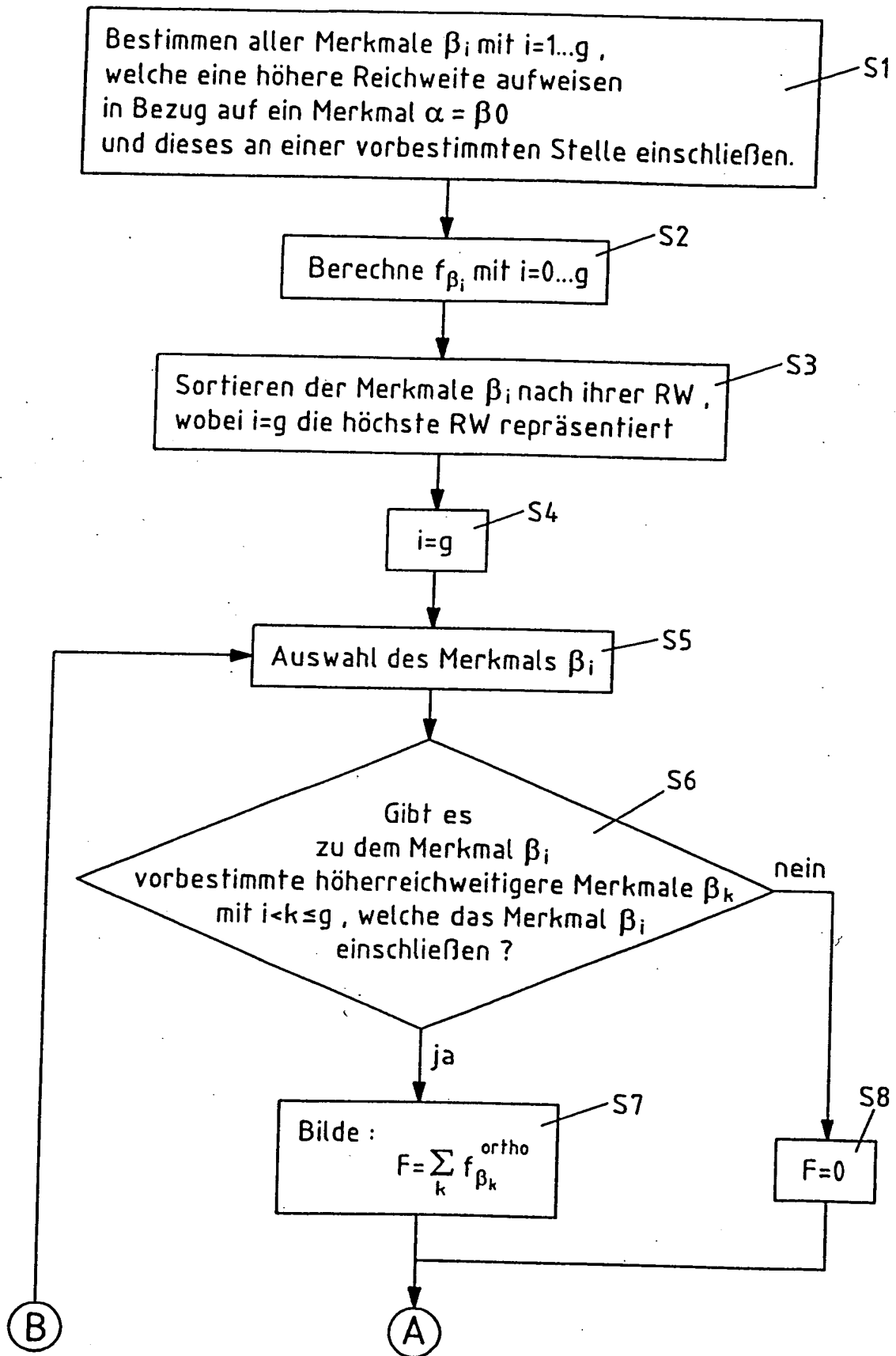


Fig.2a

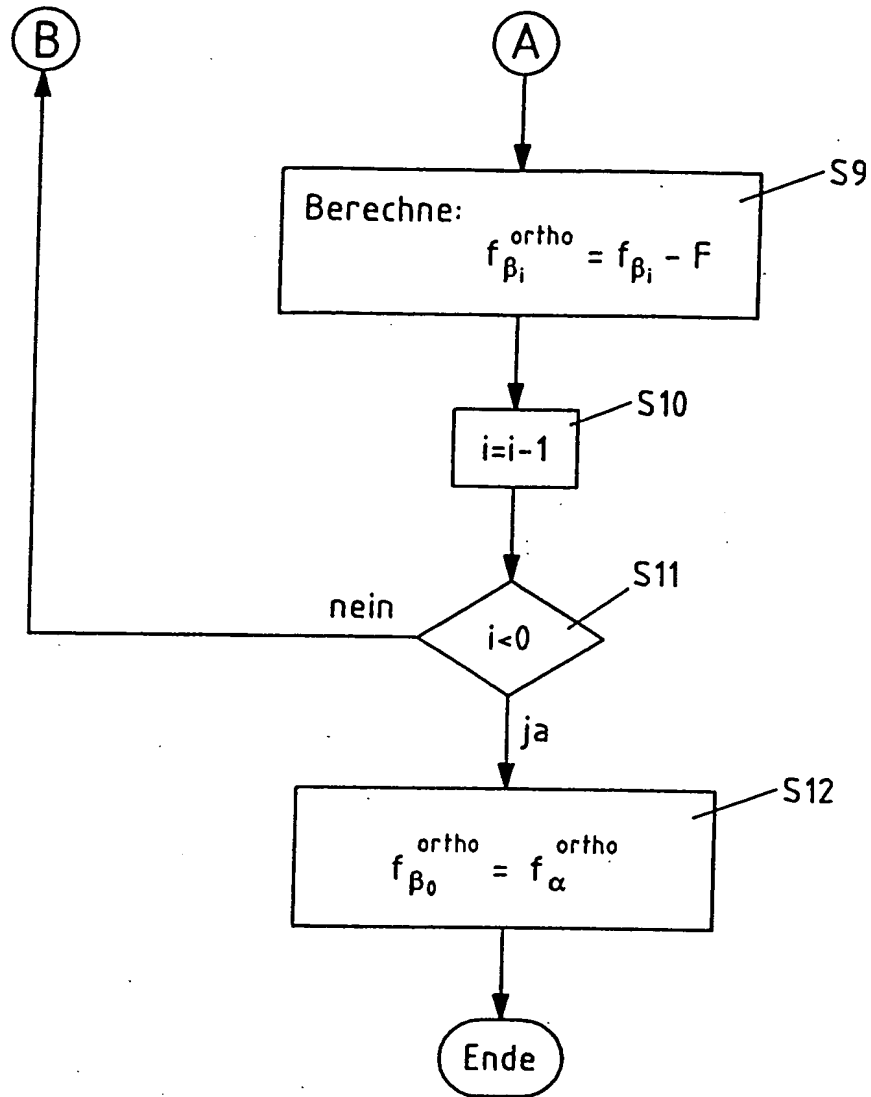


Fig.2b

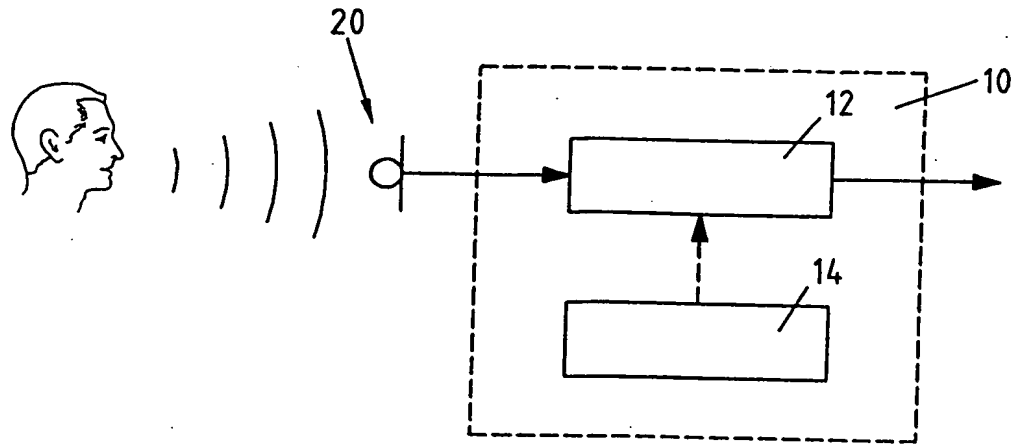


Fig.3

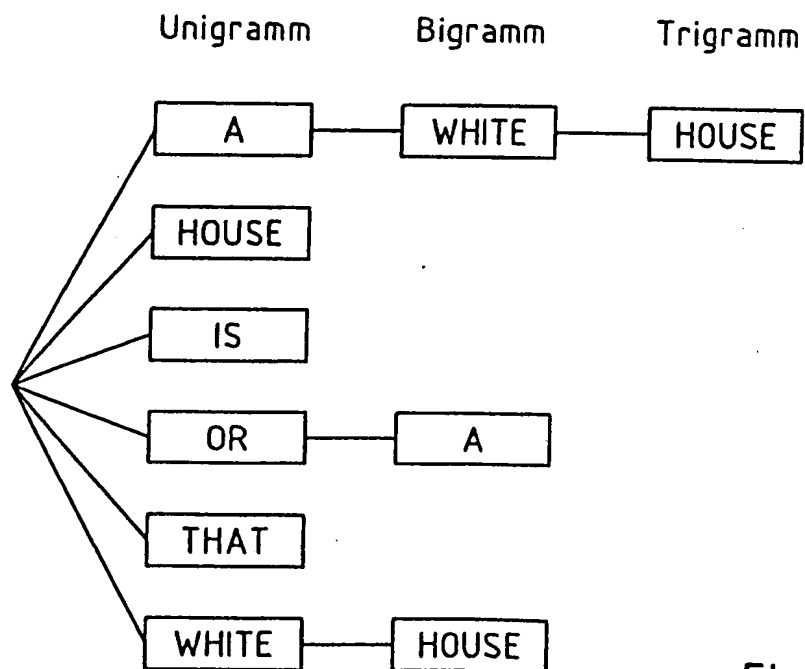


Fig.4